Brainwave: A Multidisciplinary Journal (ISSN: 2582-659X), Vol. 3, Special Issue, March 2022, pp. 85-95,

State of the Art Analysis if Covid 19 cases in the North Eastern India

Salini RoyChowdhury[1] and **Debdutta Pal**[2]

[1,2]*Brainware University,Ramkrishnapur Road,Barasat,Kolkata-700125*

Abstract

Coronavirus disease (COVID-19), has spread over the world since early 2020. The disease causes respiratory problem with manifestations, for example, cold, cough and fever. This viral disease has been declared as Pandemic on January, 2020 by International Health Regulations Emergency Committee of the World Health Organisation. India has also faced many deaths as an effect of this Corona virus. From early 2021 vaccination has started in India, as a result of vaccination the effect of this inflamm.

Keywords- COVID-19, exponential smoothing method, forecasting, decision tree.

Introduction

The author have described about several disease outbreaks that invaded humanity in World history. World Health Organization (WHO), its co-operating clinicians and various national authorities around the globe fight against these pandemics to date.The novel coronavirus appeared in the Wuhan city of China was reported to the World Health Organization (W.H.O) **[6,7,8,9,10,16]** We have a slightly different date we might count this from, and the date most scientists will recall is the day they locked their lab and went home. Although the history of Machine Learning (ML) dates back to at least the 1950s, the techniques have seen wide usage only in the last two to three decades. The main reasons are recent advances in computing power, increasing availability of open-source software, and developments in data capture and database technologies.

ML Tecniques-

Atype of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

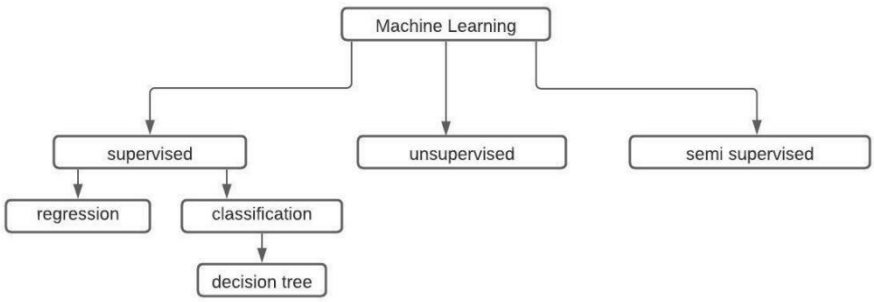

Fig1: Machine Learning Technique

Supervised Learning-

Supervised learning is based on training a data sample from data source with correct classification already assigned. Supervised machine learning techniques are applicable in numerous domains**[4,5,17]**.One standard formulation of the supervised learning task is the classification problem.

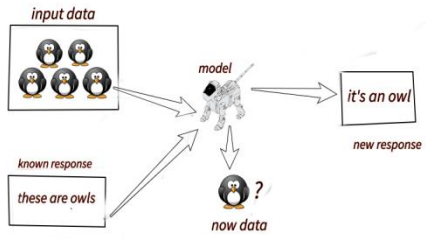

Fig2: Supervised Learning

Decision tree Algorithm-

We begin with an overview of decision trees since they are the building blocks of the SML algorithms discussed in this section.**[3,5,16]** There are many tree-based algorithms for classification and regression:

The algorithm works as follows:

1. Start from the root note with all the data.

2. Split each node into two child nodes to minimize some impurity measure (defined later). The best split is found by searching through all possible combinations of variables and their split points.

3. The tree is grown until a stopping rule is reached. Tree size is controlled by several hyper-parameters which are selected by hyper-parameter tuning.

4. Finally, data within each terminal node (or leaf) is used to prediction: node sample mean for continuous responses and majority vote or class proportions for binary/categorical responses.

Decision tree(DT) algorithm comes under Classification model which is from Supervised Learning. Decision Trees (DT) are trees that classify instances by sorting them based on feature values. Decision tree learning, used in data mining and machine learning, uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value.

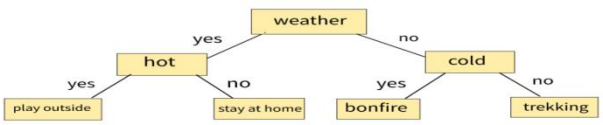

Fig3:Decision Tree

Unsupervised Learning-Unsupervised learning can be described as the general problem of extracting value from unlabelled data which exists in vast quantities. A popular framework for unsupervised learning is that of representation learning, whose goal is to use unlabelled data to learn a representation that exposes important semantic features as easily decidable factors**[12,13,16,18,20]**.The unsupervised learning algorithms learns few features from the data. Unsupervised learning We have a slightly different date we might count this from, and the date most scientists will recall is the day they locked their lab and went home.

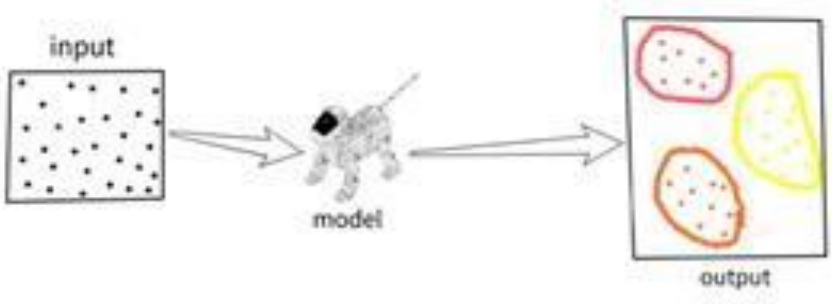

Fig4:Unsupervised Learning

Semi supervised Learning-

Semi-supervised learning considers the problem of classification when only a small subset of the observations have corresponding class labels. **[13,14,15,16,19,21,22,23]**Semi-supervised learning is a branch of machine learning that makes use of a small set of labeled data and a large set of unlabeled data to improve learning accuracy. The main downfall of this approach, it can't cluster an unknown data accurately. In the active semi-supervised learning (ASSL), the training set consists of unlabeled and labeled samples. As aforementioned, since the cost associated with the sample annotation process is high (and it can require the opinion of one or more specialists), the smallest possible set of samples should be labelled.

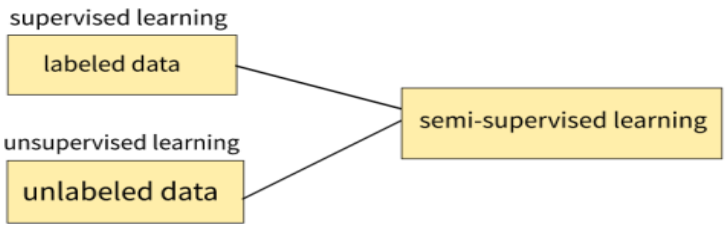

Fig 5:Semi Supervised Learning

Literature Review-

Author says, **[1]**The following preprocessing steps are applied to the data to achieve better ,accuracy, efficiency and scalability of the classification process:-

1) **Data cleaning:** This refers to the preprocessing of data with treatment of missing values**[1]**(by replacing missing value with most commonly occurring value using pandas library).

2) **Data Transformation and Reduction:** Sometimes the dataset ,may be required to be transformed or added with other datasets.

Relevance analysis:

Author says, **[1,2,3]**The dataset may contain redundant attributes. The techniques like correlation analysis can be used to find out if any two attributes are statistically related. As an example ,the high correlation between attribute A1 and A2 ,would result in removal of one of the attribute. Another relevance analysis is Attribute subset selection that finds a reduced set of attributes , such that the attained probability distribution of data classes is as near as possible to the original probability distribution using all attributes. This is how we detect attributes that do not contribute to classification. A comprehensive review is performed for the latest and most efficient approaches that have been performed by researchers in the past three years about decision trees in different areas of machine learning. Also, the details of this method, such as using algorithms/approaches, datasets, and the findings achieved are summarized. In addition, this study highlighted the most commonly used approaches and the highest accuracy methods achieved.All supervised machine learning algorithms are based on a predefined set of labels , and a training set comprised of articles which have been assigned one or more labels

Model Training:

In this phase we process the training set and we construct a classification model . This procedure includes three stages where we correlate keywords, authors and journals to one or more labels. We also record several frequency values which will be used later by the classification algorithm to effectively determine the labels of the unclassified papers. The majority of the research articles includes a set of keywords placed between the abstract and the first section.

REFERENCE	TYPE	PARAMETER	ADVANTAGE	DISADVANTAGE
[6-9]	**INTRODUCTION**			
[10]	**ML TECHNIQUES**	Atype of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.	1.With ML, you don't need to babysit your project every step ofthe way. Since it means giving machines the ability to learn, it lets them make predictions and also improvethealgorithms on their own. A common example of this is anti-virus softwares; they learn to filter new threats as they are recognized. ML is	1.Machine Learning requires massive data sets to train on, and these should be inclusive/unbiased, and of goodquality. There can also be times where they must wait for new data to be generated. 2.ML needs enoughtime to let the

			also good at recognizing spam. 2.Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. For instance, for an e-commerce website like Amazon, itserves to understand the browsing behaviors and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them.	algorithms learn and develop enough to fulfill their purpose with a considerable amount of accuracy and relevancy. It also needs massive resources to function. This can mean additional requirements of computer power for you.
[16,15,4,3,5]	**SUPERVISED LEARNING**	its use of labeled datasets to trainalgorithms that to classify data or predict outcomes accurately. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately, which occurs as part of the cross validation process.	1.The use of well-known and labelled input data makes supervised learning produce a far more accurate and reliable than unsupervised learning. With the access to labels, it can use to improve its performance on some task. 2.Efficient in finding solutions to several linear and non-linear problems such as classification, robotics, prediction and factory control. Able to solve complex problem by having hidden neuron layer	1.Performs poorly when there are non-linear relationships. One of supervised learning method like linear regression not flexible to apprehend more complex structure. It takes a lot of computation time and also difficult to append the right polynomials or interaction terms. 2.Takes a long time for the algorithm to compute by training because

				supervised learning can grow in complexity. Therefore, it is not giving result in real time since majority of world's data is unlabelled, the performance is quite limited.
[17,19,15, 12,11]	**UNSUPERVISED LEARNING**	It uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention.	1.Excels at problem where insufficient labelled dataset or identifying unknown pattern or constantly evolving. learning the concealed pattern of the data it has trained on. Makes previously unmanageable problem more solvable and more agile at finding hidden structure in past data and future prediction. 2.Simplified human task of labelling by grouping similar object and differentiating the rest. This grouped of dataset is then labelled instead of labelling it one by one.	1.Quite slow and consumes large resource memory, therefore harder to scale to larger datasets. Moreover, it only presumes the underlying clusters in the dataset are glob-shaped. 2.The outcomes are not that accurate due to it is mostly about prediction. In addition, we do not know the number of classes, therefore the results are not certain. Unsupervised learning is less adept to solve narrowly defined problem
[18,21,22, 15,14,12, 20]	**SEMI-SUPERVISED LEARNING**	It is a combination of supervised and unsupervised machine learning **methods**. ... In semi-supervised learning, an algorithm learns from a dataset that includes both labeled and unlabeled data, usually mostly unlabeled.	1.The learning agent or system themselves, crafts the data on its own by interacting with the environment. Does not require a huge amount of data to train itself to develop a generalized	1.Defining the reward isdifficult. It is usually given or hand-tuned by the algorithm designer. Reward function must adhere to the exact goal or risk overfitting

			formula like supervisedlearning 2.Is one of the nearest to thetype of learning that humans and mammals do. In fact, majority of the fundamental algorithm of RL are derived from human brain and neurological system	and also stranded at local optima. 2.Need a lot of training data and need some time to train to be more accurate and efficient compared to other learning algorithm.

STEPS TO BE FOLLOWED-

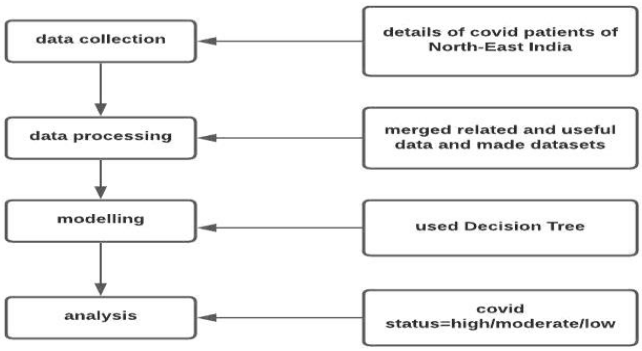

Fig6:Steps To Be Followed

PROPOSED ALGORITHM-

Total case (T_{Test}): It defines the total number of Tests that include Delta (T_{Delta}) and (T_{delta+}}that have occurred in one month for Covid19 ina specific state.

Step 1 : start

Step 2 : check the parameters(real values) of the attributes

Step 3 : determine class to make decision

Step 4 : ifP_{ratio}>=0.1 then // P_{ratio}in equation (iii)

Covid_status:=high

Step 5 : if P_{ratio}<0.1 &&P_{ratio}>=0.5&&U_{ratio}>.05 then

Covid_status:= moderate

Step 6 : if P_{ratio}<0.5 && U_{ratio}<.05 then

Covid_status:= low

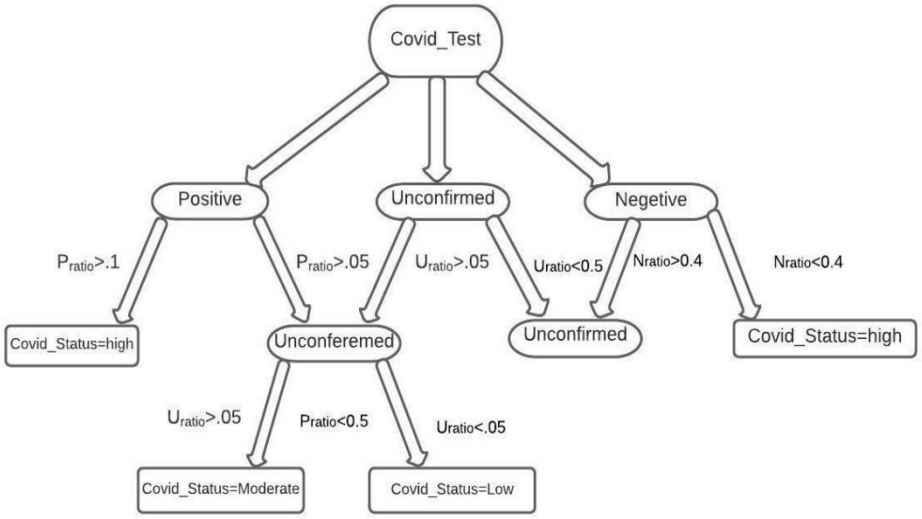

Fig7:Decision Tree

DATASET-Here we have fetched datasetsof North-East states of India-Assam,Nagaland,Odisha,Tripura and West Bengal. As an example from all the taken states in Table 1,dataset of Tripura is shown-

No.	1: month String	2: total_tested Numeric	3: positive Numeric	4: p_ratio Numeric	5: negative Numeric	6: n_ratio Numeric	7: unconfirmed Numeric	8: u_ratio Numeric	9: **decission** Nominal
1	JAN	292919.0	3093.0	289628.0	0.988764...	7770.0	0.026526104	0.02084...	low
2	FEB	1130127.0	23074.0	1107053.0	0.979582...	7511.0	0.006646156	0.04069...	low
3	MAR	2741230.0	68037.0	2456818.0	0.896246...	8288.0	0.00302346	0.05159...	moderate
4	APR	6263584.0	201636.0	6061948.0	0.967808...	8029.0	0.001281854	0.06434...	moderate
5	MAY	1.0231761E7	530848.0	9665403.0	0.944647...	8288.0	8.10027E-4	0.10390...	high
6	JUN	1.299398E7	820027.0	1.21146...	0.932329...	8029.0	6.17902E-4	0.12646...	high
7	JUL	1.5158117E7	891189.0	1.41739...	0.935073...	8288.0	5.4677E-4	0.11791...	high
8	AUG	1.7790528E7	926419.0	1.67330...	0.940559...	8288.0	4.65866E-4	0.10450...	high
9	SEP	1.7910388E7	900706.0	1.69108...	0.944191...	8029.0	4.48287E-4	0.10083...	high
10	OCT	1.9842632E7	1034920.0	1.87740...	0.946148...	8288.0	4.17687E-4	0.10437...	high
11	NOV	2.0260196E7	1047167.0	1.92130...	0.948314...	8029.0	3.96294E-4	0.10335...	high
12	DEC	2.4995688E7	1275074.0	2.37104...	0.948580...	8288.0	3.31577E-4	0.10202...	high
13	JA	2.8386604E7	1488855.0	2.68980...	0.947559...	6993.0	2.46349E-4	0.10488...	high

Table1: Dataset of Tripura

Performance Analysis-

We have used WEKA to show the performance-

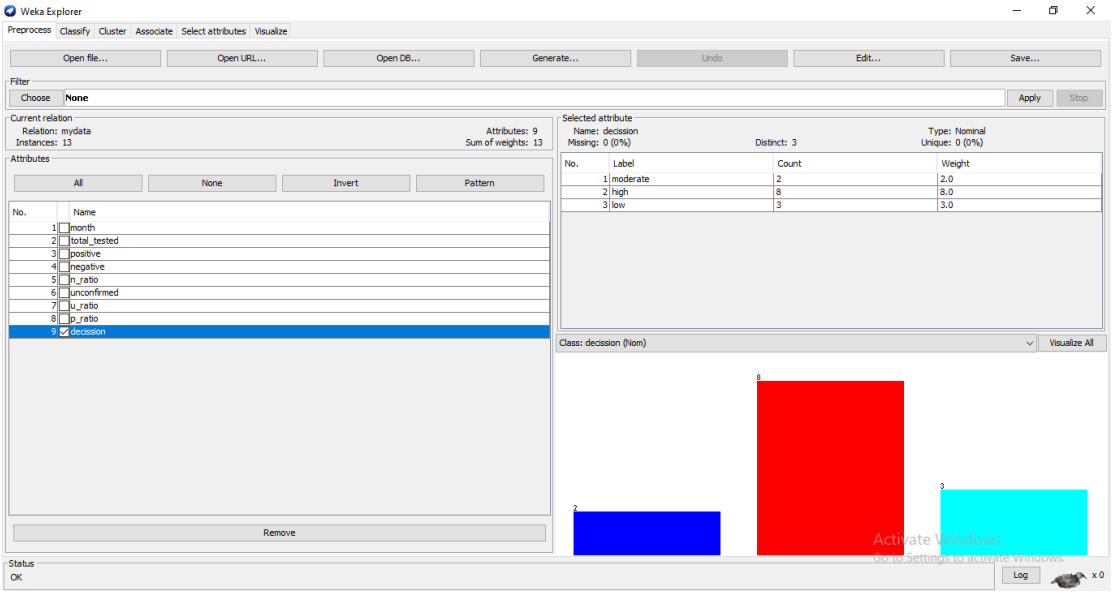

Fig8:The preprocess tab of Weka is showing the status of Covid19 with parameters of high,moderate and low.

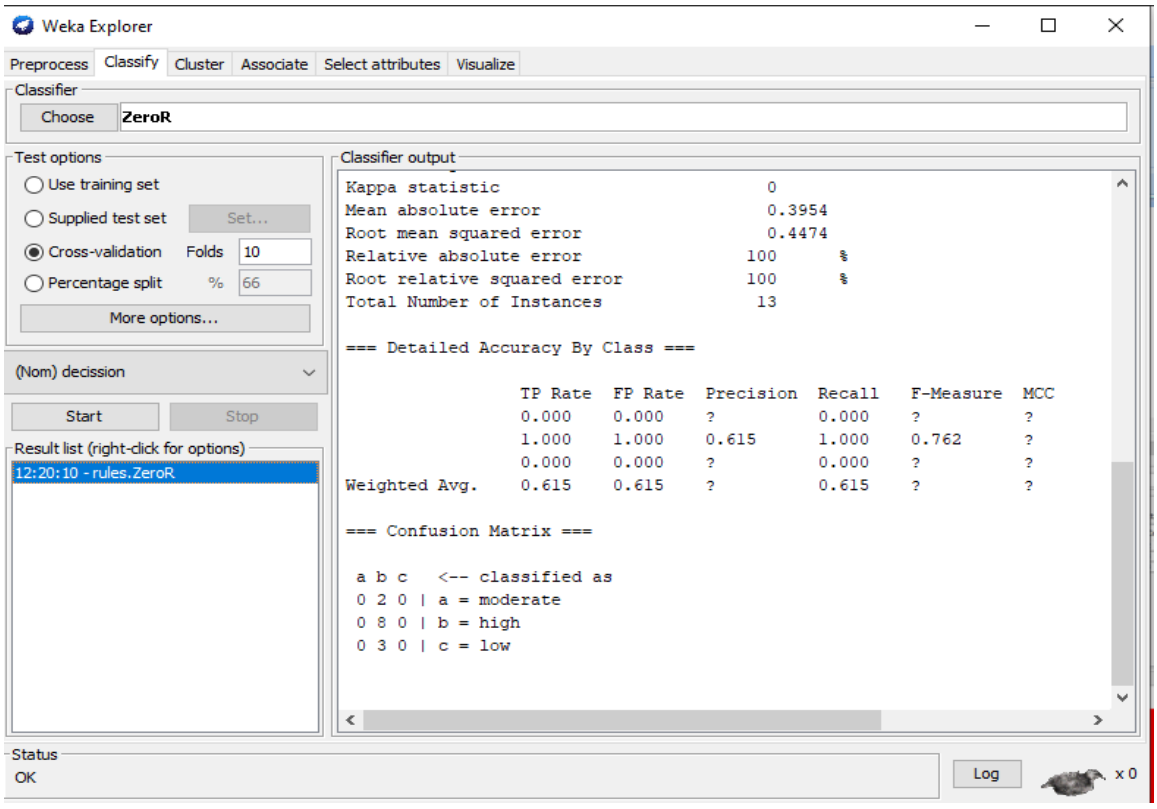

Fig9:classified as low,moderate,high

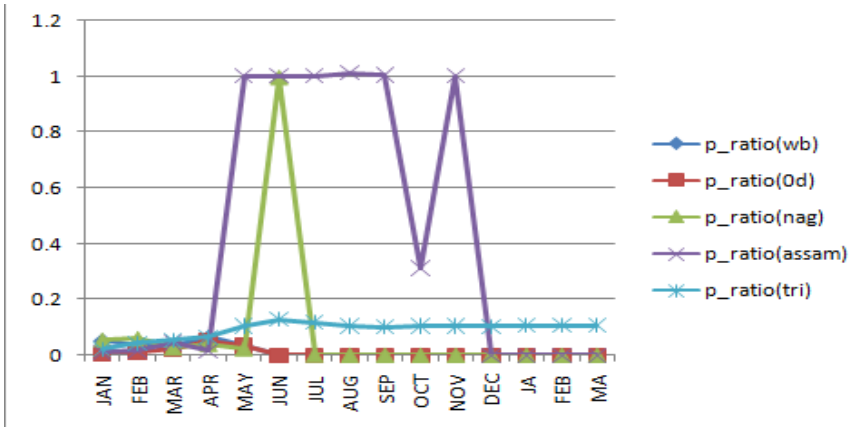

Fig10:Observation of the dataset of Tripura

Conclusion-

The covid-19 epidemic has presented many countries in the world with an unprecedented public health crisis now days. The effect of covid-19 has significant role in world economy. Using Machine learning, a model can be invented to predict and analyze the effect of epidemic in public health with the course of time. Information and communication technology support in the decision-making process based on the previously collected data. As the size of the collected data is huge that make it difficult

Refrences-

1.Archit Verma1,"STUDY AND EVALUATION OF CLASSIFICATION ALGORITHMS IN DATA MINING",Aug,2018

2.Leonidas Akritidis, Panayiotis Bozanis, "A Supervised Machine Learning Classification Algorithm for Research Articles"-Proceedings of the 28th Annual ACM Symposium on Applied Computing, Pages
115–120, 2013
3.BahzadTaha Jijo1*, Adnan Mohsin Abdulazeez2," Classification Based on Decision Tree Algorithm for Machine Learning"Jan,2021

4.ArirunaDasgupta, AsokeNath," Classification of Machine Learning Algorithms",March,2016

5.Osisanwo F.Y.*1 ,Akinsola J.E.T.*2, Awodele O.*3 , Hinmikaiye J. O.*4 , Olakanmi O.*5 ,Akinjobi J. **6"Supervised Machine Learning Algorithms: Classification and Comparison ",June,2017

6.JOURNAL OF BIOLOGICAL RHYTHMS, Vol. 36 No. 1, February 2021 3 DOI:10.1177/0748730421993352 © 2021 The Author(s) Article reuse guidelines: sagepub.com/journals-permission-h. Nicolas Cermakian Douglas Research Centre, McGill University Mary E. Harrington Neuroscience Program, Smith College

7. Nicholas Grubic1 ,Shaylea Badovinac2 and Amer M Johri3,"Student mental health in the midst of the COVID-19 pandemic"-A call for further research and immediate solutions,pages517-518,2020

8.AkibMohiUd Din Khanday1 • Syed Tanzeel Rabani1 • QamarRayees Khan1 •Nusrat Rouf1 • MasaratMohiUd Din2,"Machine learning based approaches for detecting COVID-19 using clinical text data",June 30,2020

9. Samuel Lalmuanawma , Jamal Hussain , LalrinfelaChhakchhuak,"Applications of Machine Learning and Artificial Intelligence for Covid-19 (SARS-CoV-2) pandemic:"-A review,2020

10.KonstantinaKouroua , Themis P. Exarchosa,b , Konstantinos P. Exarchos a , Michalis V. Karamouzis c , Dimitrios I. Fotiadis a,b,"Machine learning applications incancer prognosis and prediction" ,pages 8-17,2015

11.MemoonaKhanuma ,TahiraMahboob b ,"A Survey on Unsupervised Machine Learning Algorithms for Automation, Classification and Maintenance",June 13,2015

12.AyonDey ,"Machine Learning Algorithms: A Review,pages 1174-1179,2016

13. Y C A Padmanabha Reddy, ViswanathPulabaigari, Eswara Reddy B,"Semi-supervised learning: a

brief review",Feb,2018

14.XiaohuaZhai∗ ,Avital Oliver∗ Alexander Kolesnikov∗ , Lucas Beyer∗ ,"S 4L: Self-Supervised Semi-Supervised Learning"

15.Linwei Hu, Jie Chen, Joel Vaughan, Hanyu Yang, Kelly Wang, AgusSudjianto, and Vijayan N. Nair1,"Supervised Machine Learning Techniques: An Overview with Applications to Banking",July 26,2020

16. R. Sathya ,AnnammaAbraham,"Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification",2013

17. Xi Chen†‡, Yan Duan†‡, Rein Houthooft†‡, John Schulman†‡, IlyaSutskever‡ , Pieter Abbeel†,

"InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets"

18.Diederik P. Kingma∗ ,J. DaniloRezende† , Shakir Mohamed†,Max Welling ,"Semi-supervised Learning with Deep Generative Models"

19.Jenq-HaurHaur Wang, Hsin-Yang Wang, Yen-Lin Chen, Chuan-Ming Liu,"A constructive algorithm for unsupervised learning with incremental neural network", pages 188-196(2015)

20.GuilhermeCamargo1 , Pedro H. Bugatti1 , Priscila T. M. SaitoID1,2*,"Active semi-supervised learning for biological data classification" , August 19, 2020

21. Alexander Ligtharta ,CagatayCatal b,BedirTekinerdogan a ,"Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classification", 25 December 2020.

22. Federica Bisio, Paolo Gastaldo, and Rodolfo ZuninoDITEN,"Semi-supervised machine learning approach for unknown malicious software detection",2014